

# Reconocimiento de Patrones

Version 2021-2

## Selección de Características

[ Capítulo 3 ]

**Dr. José Ramón Iglesias**

DSP-ASIC BUILDER GROUP

Director Semillero TRIAC

Ingeniería Electronica

Universidad Popular del Cesar

# 5 reasons why to select features

1. To avoid non-discriminative features
2. To avoid correlated features
3. To simplify the testing stage
4. To avoid false correlations
5. To avoid the curse of dimensionality

# Example: Classification of Coins

A

B



# 5 reasons why to select features

1. To avoid non-discriminative features

1. To avoid non-discriminative features - ~~Grayvalue~~ average ☹



# 5 reasons why to select features

1. To avoid non-discriminative features
2. To avoid correlated features

2. To avoid correlated features - Area and ~~Diameter~~ ☹



# 5 reasons why to select features

1. To avoid non-discriminative features
2. To avoid correlated features
3. To simplify the testing stage



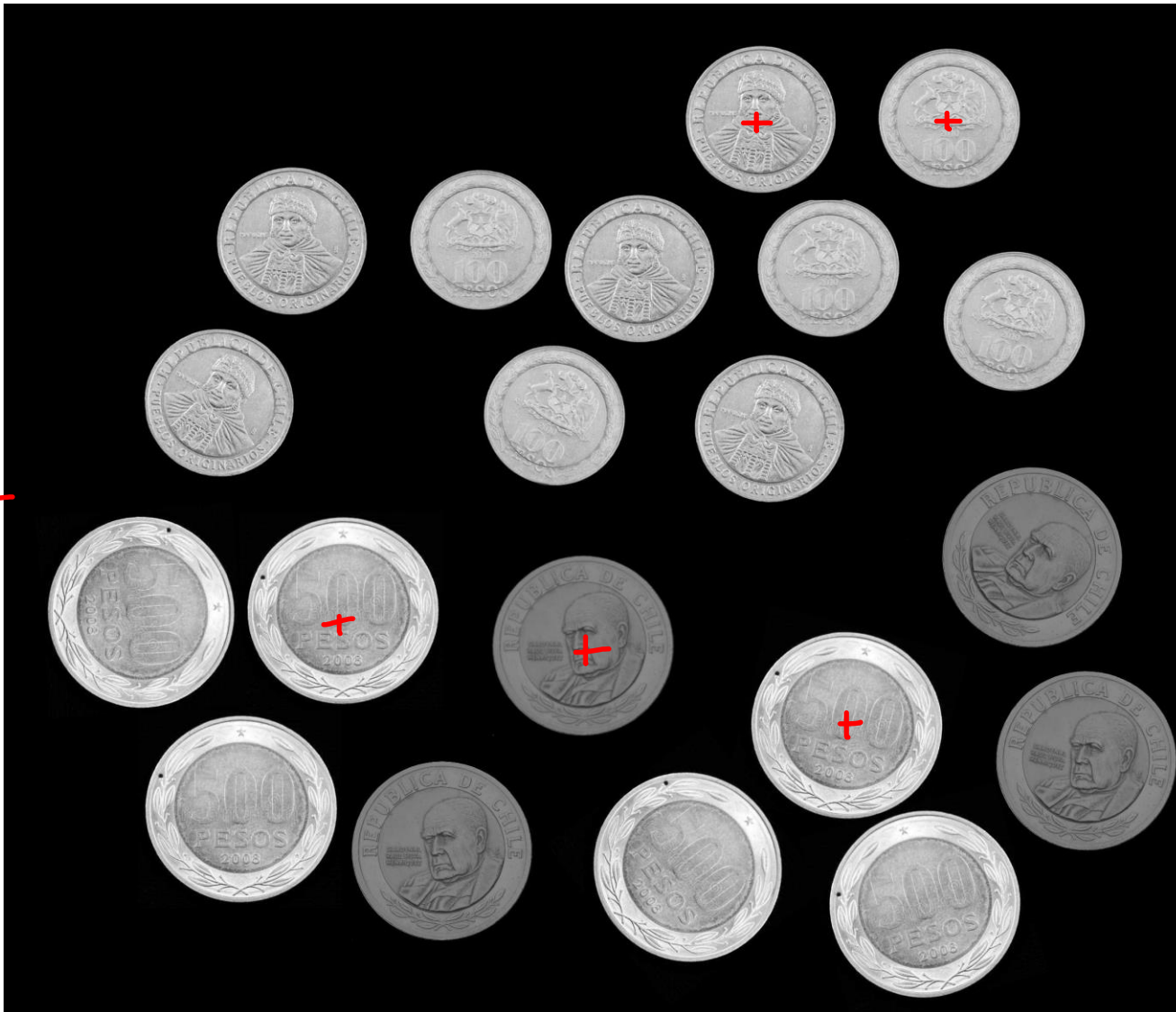
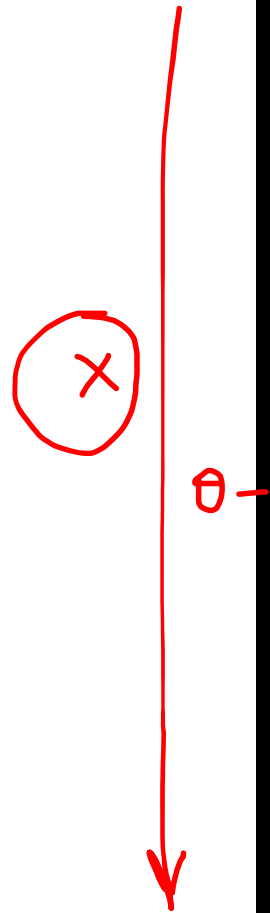
### 3. To simplify the testing stage - Area vs. ~~Hu+Flusser~~



# 5 reasons why to select features

1. To avoid non-discriminative features
2. To avoid correlated features
3. To simplify the testing stage
4. To avoid false correlations

#### 4. To avoid false correlations - ~~Location~~ ☹



# 5 reasons why to select features

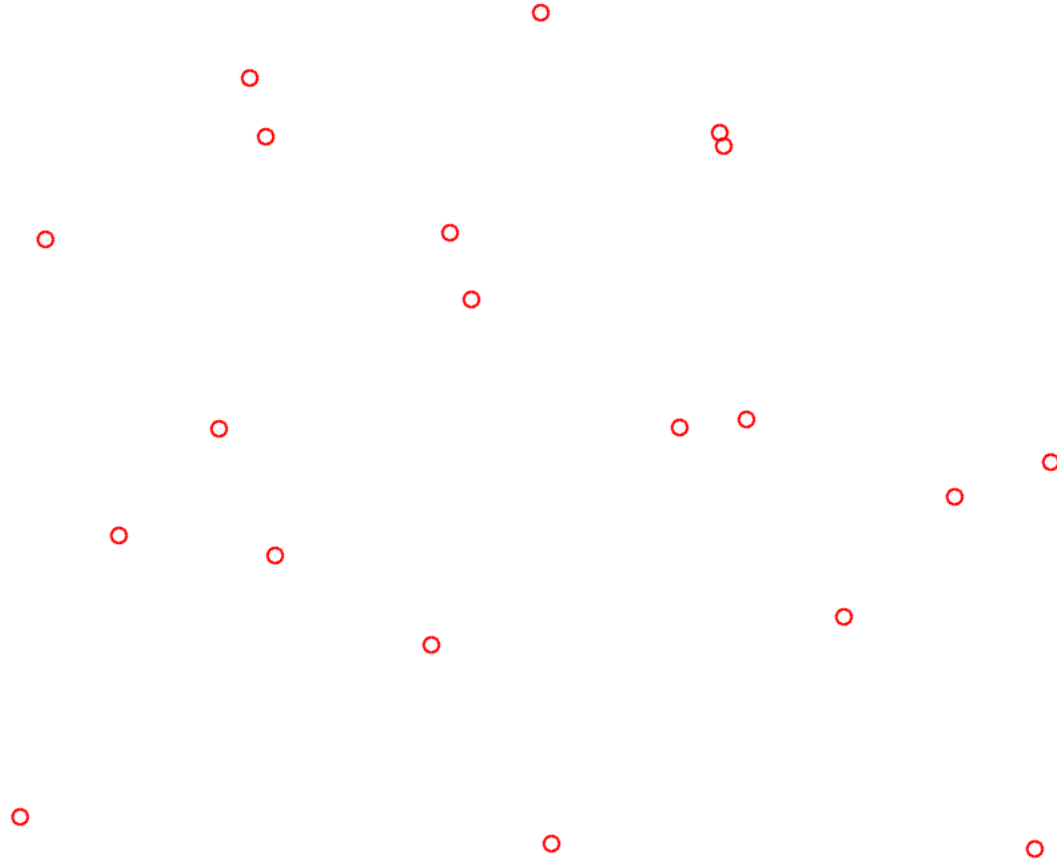
1. To avoid non-discriminative features
2. To avoid correlated features
3. To simplify the testing stage
4. To avoid false correlations
5. To avoid the curse of dimensionality

# Curse of dimensionality - 1D



(20 points in 1D)

# Curse of dimensionality - 2D

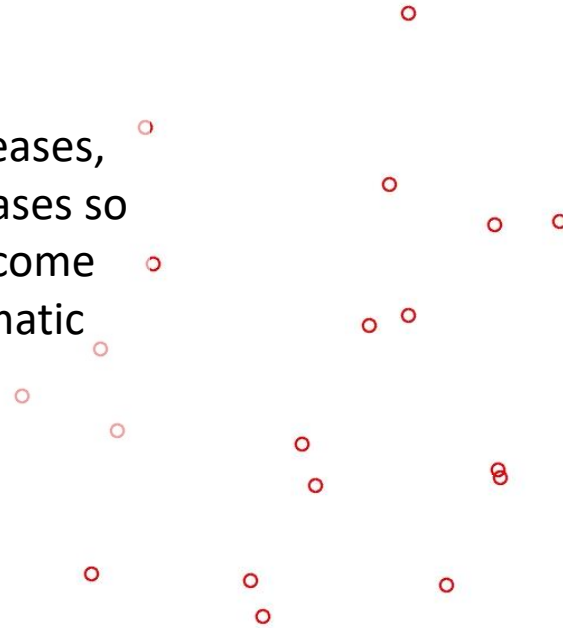


(20 points in 2D)

# Curse of dimensionality - 3D

When the dimensionality increases, the volume of the space increases so fast that the available data become sparse. This sparsity is problematic for any method that requires statistical significance.

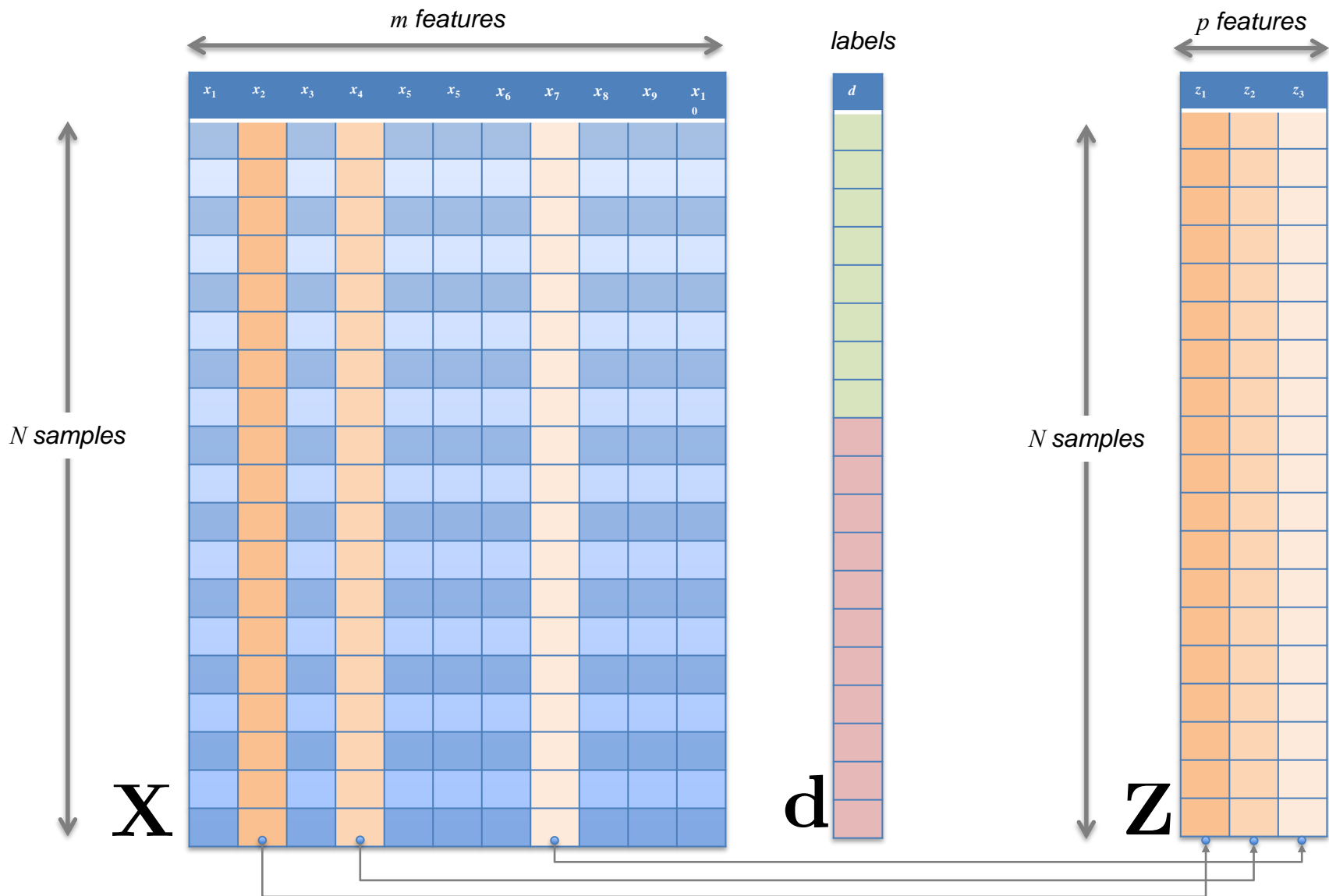
[Wikipedia]



(20 points in 3D)

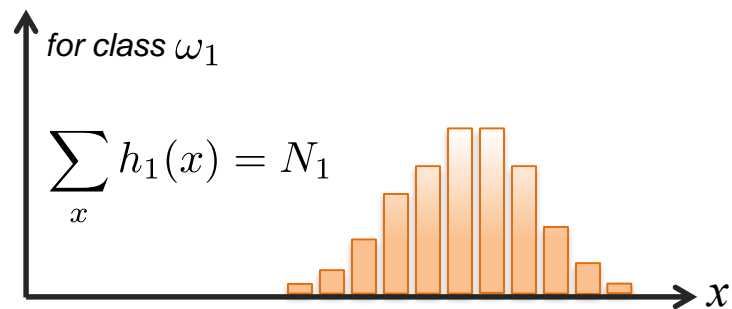
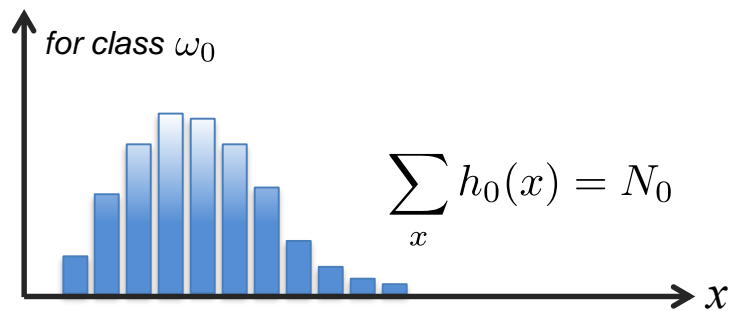
*Extracted features*

*Selected features*

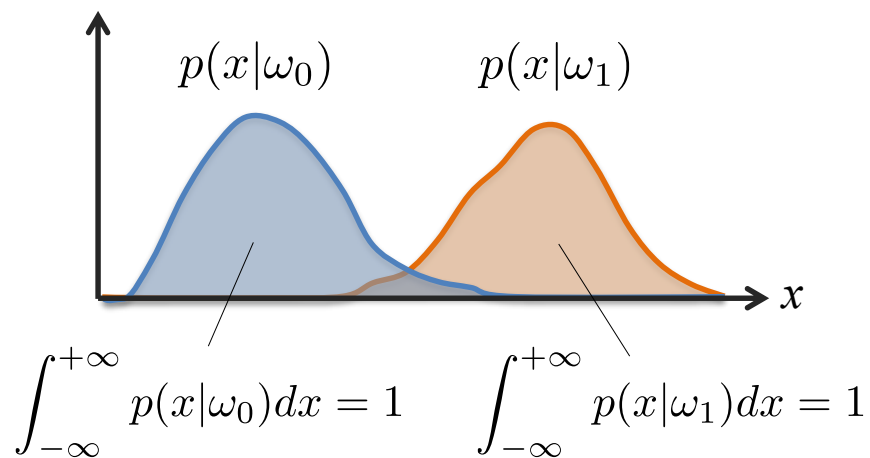




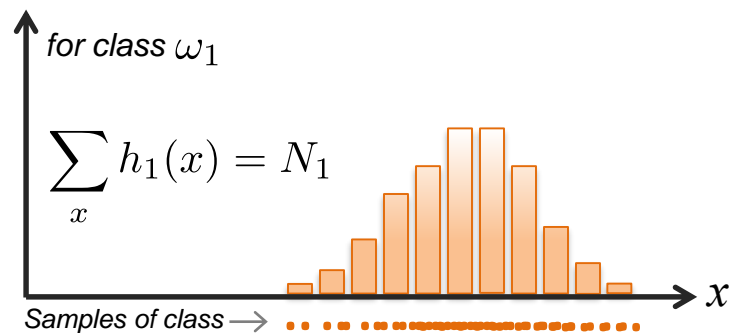
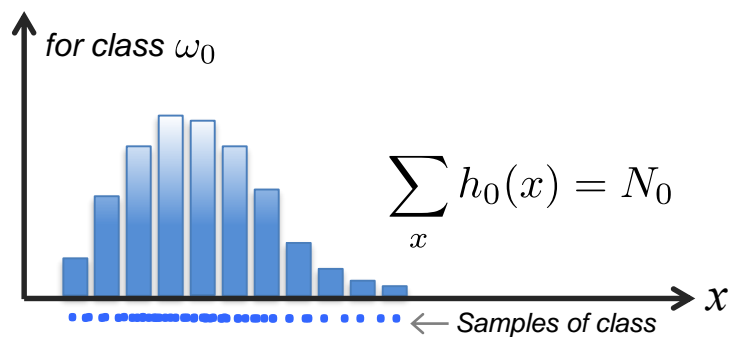
# Frequency distributions



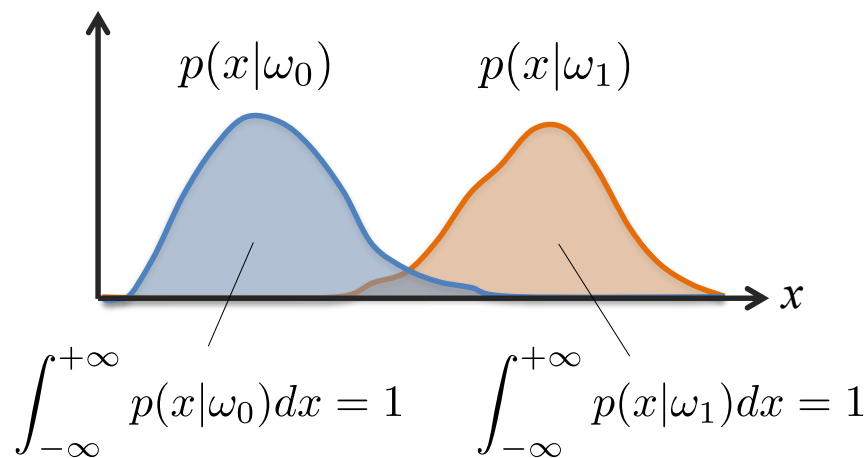
# Probability density functions

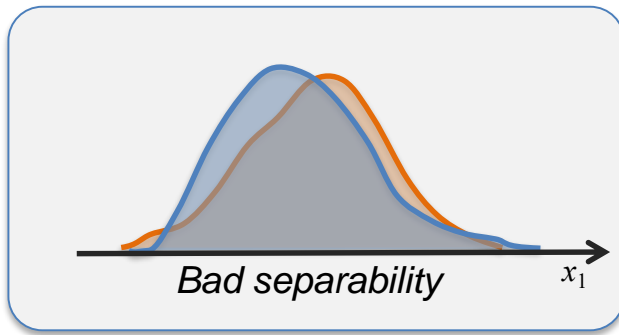


# Frequency distributions

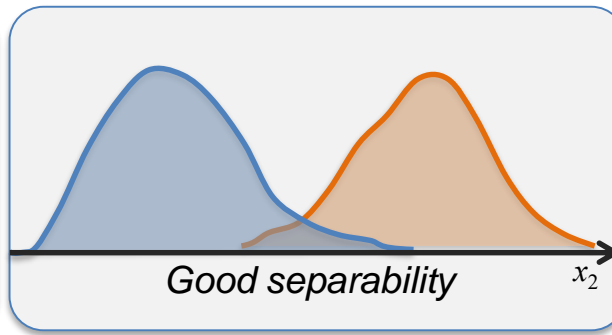


# Probability density functions

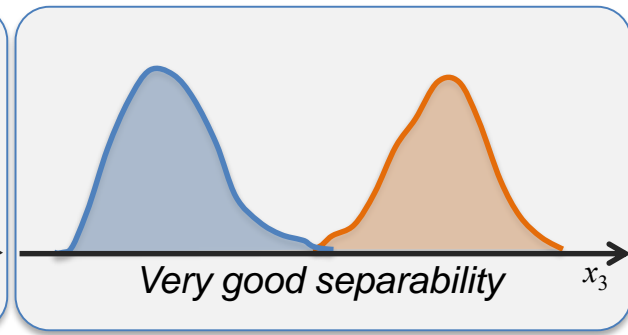




score = LOW



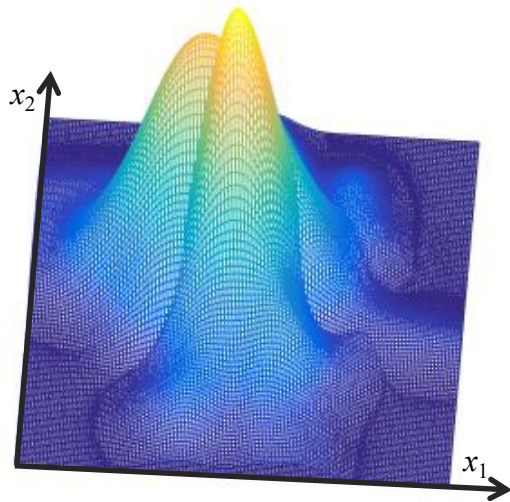
score = MEDIUM



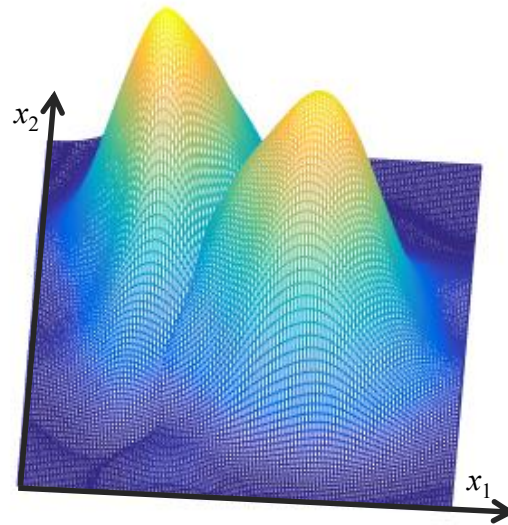
score = HIGH



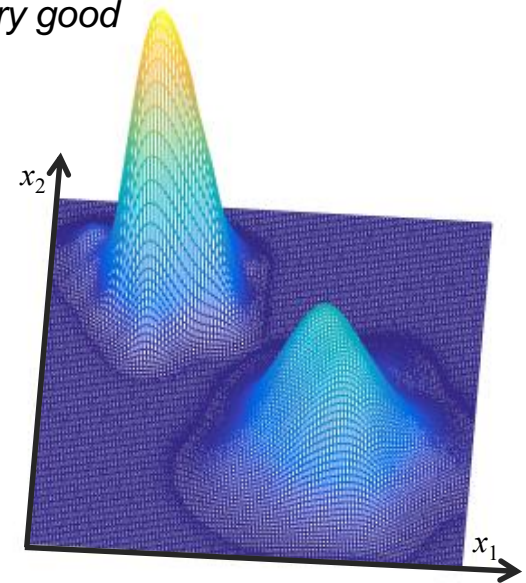
*Bad*



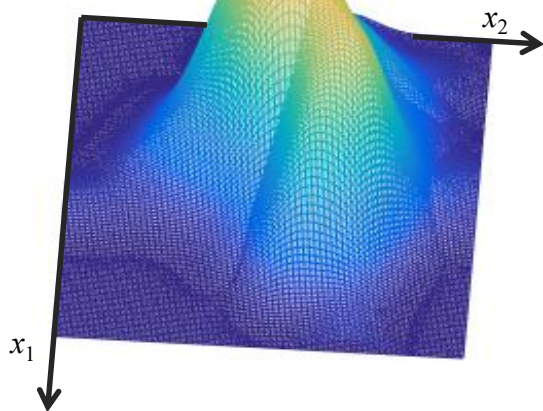
*Good*



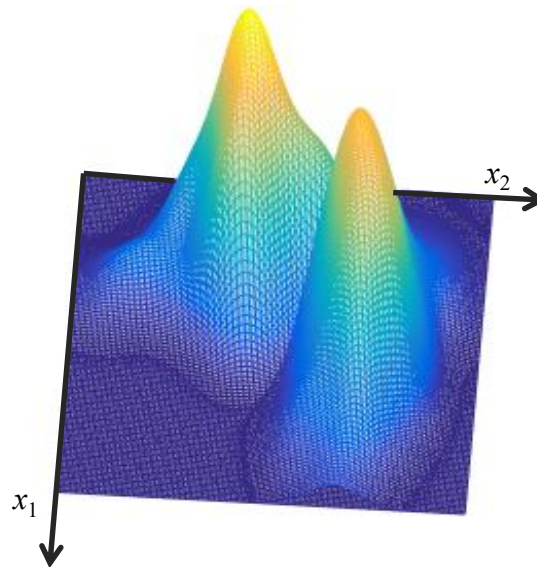
*Very good*



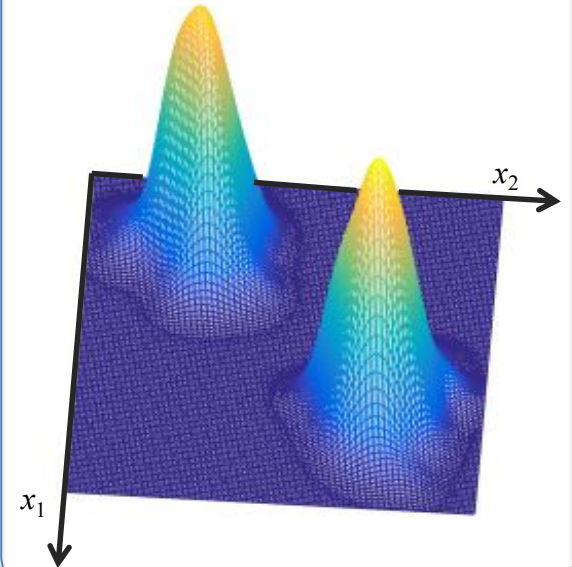
*Bad*



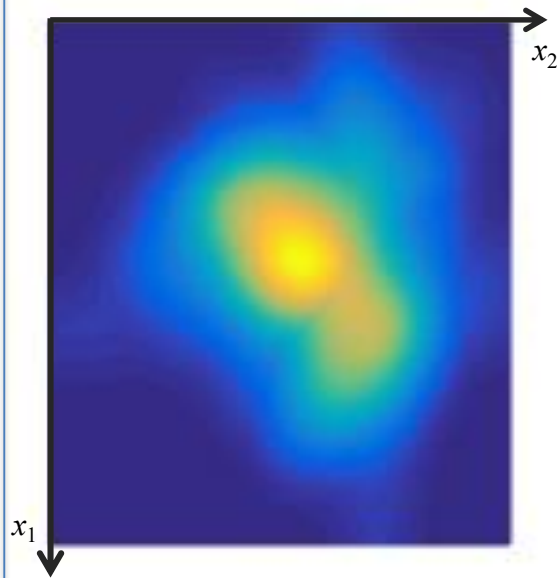
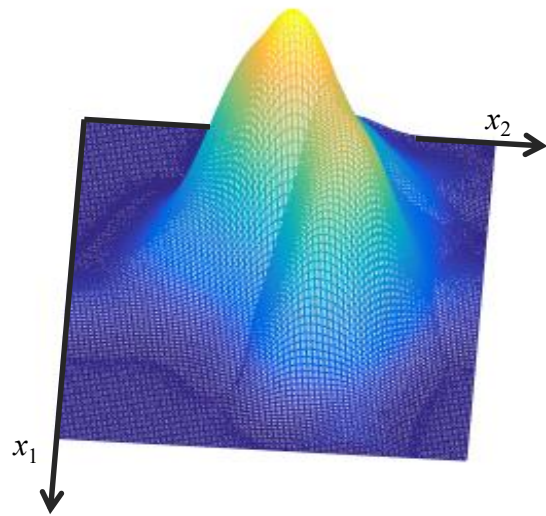
*Good*



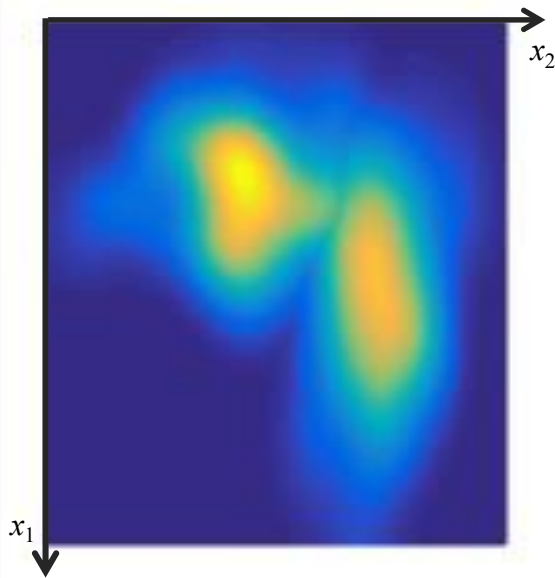
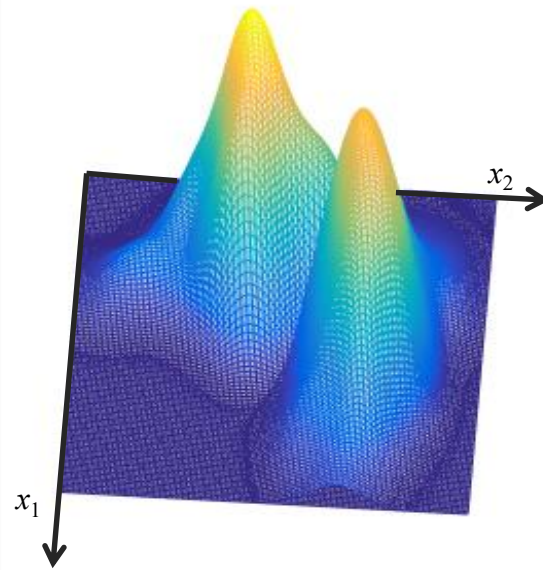
*Very good*



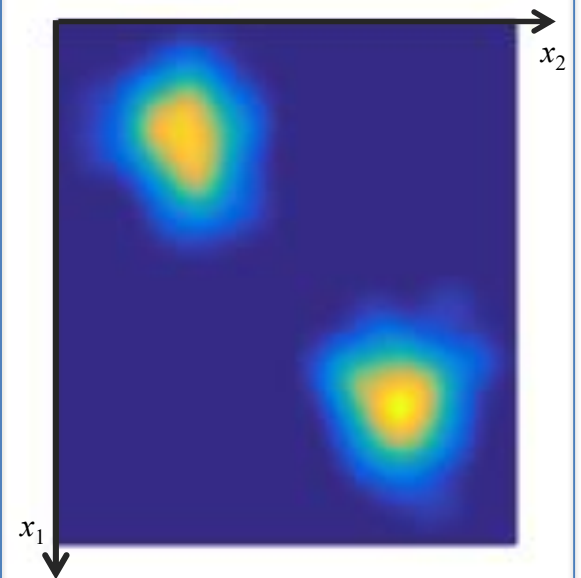
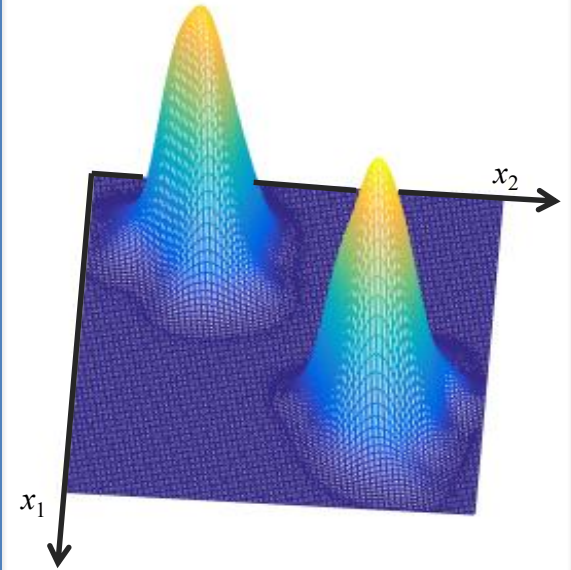
*Bad*



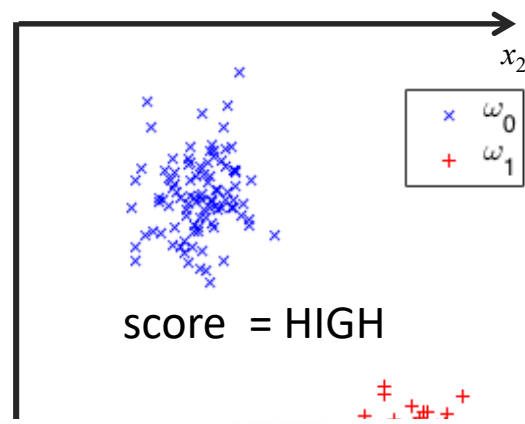
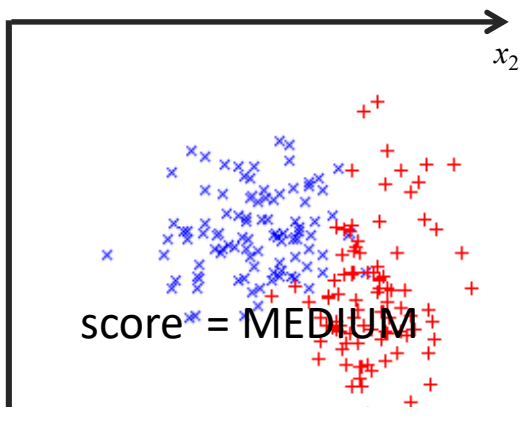
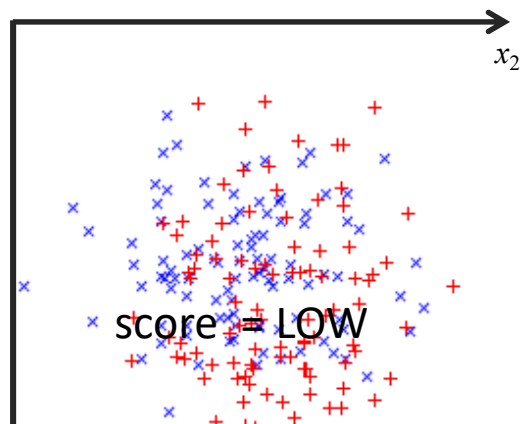
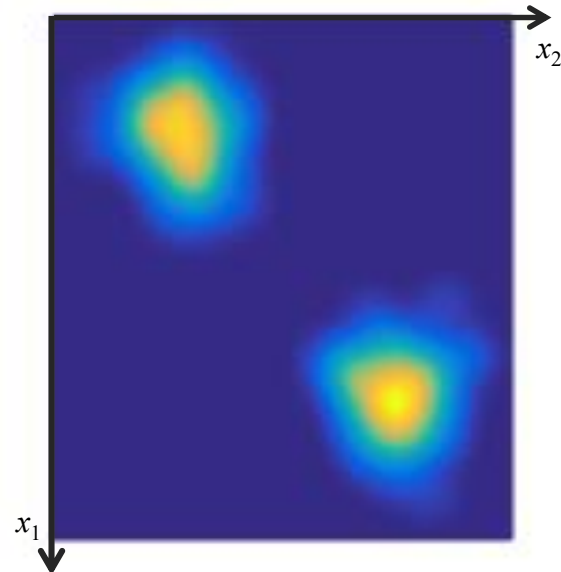
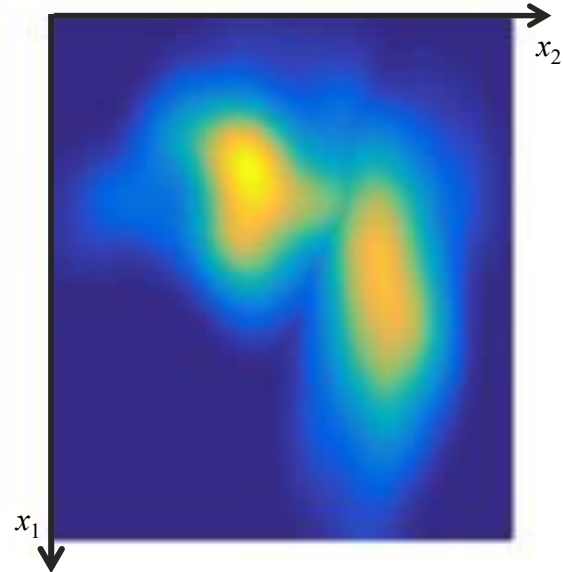
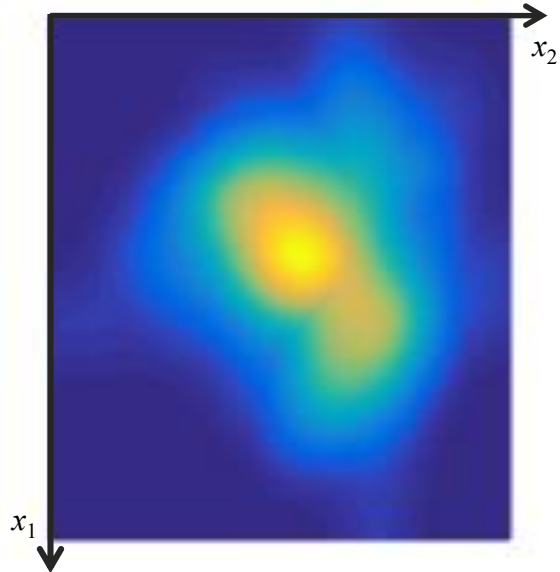
*Good*



*Very good*







*Selected features must be...*

*Can our objects of interest be everywhere?*

*yes* →

*invariant to translation*

*Can our objects of interest be in any orientation?*

*yes* →

*invariant to rotation*

*Can our objects of interest be of any size?*

*yes* →

*invariant to scale*